Biomimetic Carbohydrate Recognition

Anthony P. Davis

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Carbohydrates are important but challenging targets for supramolecular chemists. They possess complex, irregular and variable structures, and are strongly attracted to water, their natural environment. This tutorial review describes work on synthetic receptors which bind carbohydrates through non-covalent interactions, mimicking the strategies used in biology. Emphasis is placed on systems which operate in purely aqueous solution, without involvement of organic solvents. Although the problem is difficult, the careful design of complementary cavities can lead to surprisingly good results. In particular, a receptor for glucose has achieved performance which generally matches biology, and augurs well for real-world applications.

Key learning points

(1) Principles of carbohydrate recognition through non-covalent interactions; requirement for full host-guest complementarity, especially in water.

(2) Successful approaches using preorganised hydrogen bonding groups in organic solvents.

(3) Receptors for operation in aqueous solution, especially the "temple" family for all-equatorial substrates.

(4) High levels of complementarity are possible, at least for glucose, leading to remakable affinity and selectivity.

School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK. E-mail: Anthony.Davis@bristol.ac.uk

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

1. Introduction

The design of carbohydrate receptors is a major challenge for supramolecular chemists. Carbohydrates are clearly important substrates. They are the most abundant molecules in biology, and are highly versatile. They serve as fuels (e.g. starch), building materials (e.g. cellulose), and form the backbone for the genetic code (ribose in DNA). They show exceptional structural variability, which is widely exploited in labelling systems for cells and proteins.[1,2] Agents which bind to saccharides could serve many medical purposes. Perhaps the most obvious relate to diabetes, where selective glucose receptors could be exploited in blood glucose monitors[3] and glucose-responsive insulin.[4] However, there are many other possibilities, including synthetic antibodies to cancer cells, anti-infectives, anti-inflammatories, and many diagnostic applications.[1,2,5]

From a theoretical viewpoint, carbohydrate recognition is representative of a wider issue – how do we bind polar molecules in water, where solvent already binds strongly to both substrate and complementary (polar) binding site? The problem is especially acute for saccharides, where hydroxyl and ether groups are predominant and most examples are quite similar to clusters of water molecules. Distinguishing these "hydromimetic" targets from solvent seems an almost impossible task. Indeed, biological carbohydrate recognition is generally quite weak. Lectins, the major class of carbohydrate-binding proteins, often bind monosaccharides with $K_a < 10^3$ M^{-1}.[6] Nonetheless, higher affinities are possible, especially for the bacterial periplasmic proteins which help guide the organisms towards food sources.[7] Biology thus encourages and focuses our research. It confirms that strong binding is feasible, while posing questions about the principles which underlie success.

Research into carbohydrate receptors has a considerable history, going back at least to the 1980s. Two main approaches have been taken. One exploits the reversible reaction between boronic acids and diols to give cyclic boronates, allowing the design of binding agents which are simple, accessible, and effective in the natural medium of water. These systems have been covered in depth by others[2,3] and, as they are not biomimetic, fall outside the present scope. The alternative relies on noncovalent bonding and more closely mimics the strategy used by biology.[5,8] This article will discuss how receptors based on this approach have been developed, starting in unnatural media, moving to truly biomimetic systems which operate in aqueous solution, and finally to a system which both mimics and matches biology. Particular emphasis is placed on water as solvent, complementing earlier accounts which provide good coverage of organic media.[5,9] The progression illustrates the potential of supramolecular chemistry to achieve quite surprising goals, given sufficient focus and persistence.

2. General Principles

As illustrated in Fig. 1, carbohydrates are polyols with occasional presence of other polar functionality (e.g. NHAc, CO_2H). The monosaccharide units are cyclic, possessing distinct conformations, so the structures tend to be geometrically well-defined. In terms of noncovalent interactions, the polar groups can act as H-bond donors or acceptors, the oxygen atoms can serve as electron pair donors to metal ions, and the CH groups in unfunctionalised areas can also act as weak H-bond donors. Reducing sugars, with C1-OH, occur as a mixture of anomers (illustrated for glucose **1**).

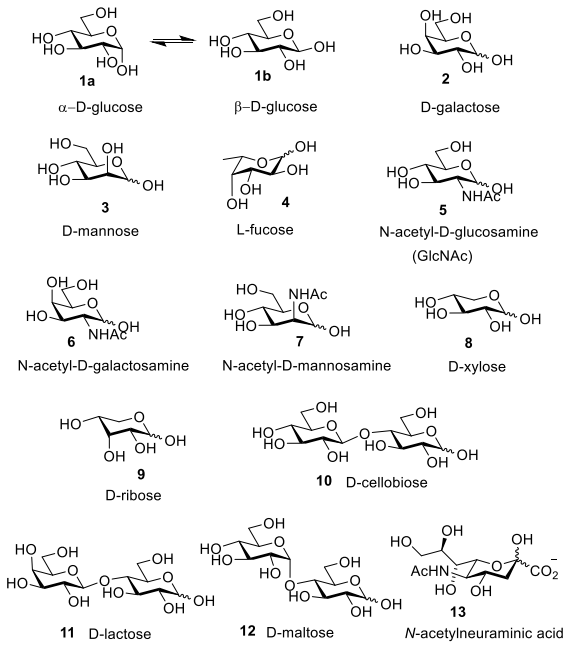


Fig. 1 Selected biologically important carbohydrates.

With so many polar groups, carbohydrates are quite easy targets in non-polar organic media. A solvent such as chloroform does not depress interactions such as H-bonding, so preorganised arrangements of polar binding groups should be effective (Fig. 2a). As implied by Fig. 2a, complementarity with apolar surfaces may not be important. The interactions between scaffold and substrate CH are likely to be weak, and not so different from contacts with solvent, so the exact framework design may be less critical.

For carbohydrate recognition in water, the situation is very different. When a saccharide -OH binds to a polar group in a receptor it must displace a water molecule, so the process is nearly isoenergetic. Polar interactions should not be worthless because an array of H-bonding groups which perfectly complements the target should not be ideal for solvent water. Nonetheless, they are severely weakened. On the other hand, compensation is available if the apolar CH groups in the carbohydrate make contact with apolar

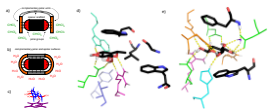


Fig. 2 (a) Design strategy for carbohydrate recognition in an organic solvent such as chloroform, focusing on complementarity between polar binding groups. (b) Moving to water, where matching apolar surfaces may be equally important. (c) CH-π interactions between an aromatic surface and axial carbohydrate CH groups. (d) X-ray crystal structure of β-D-fucose in the binding site of *Aleuria aurantia* lectin (see ref. 11). Residues contributing to apolar interactions are shown as thicker tubes with standard colouring. Selected polar hydrogens are shown, the rest are concealed. (e) Crystal structure of β-D-glucose complexed to the periplasmic galactose-binding protein from *E. coli* (see ref. 7). Display conventions as for (d).

surfaces in the receptor. "High energy" water molecules are displaced from both and are free to move into bulk, lowering the free energy. To take advantage of this, the receptor must be fully complementary to the substrate, matching both polar and apolar moieties (Fig. 2b). There is particular benefit if the receptor's apolar surfaces are aromatic, allowing the formation of CH-π bonds (Fig. 2c). Although it is not obvious that this interaction will translate to water (OH-π bonds are also possible), there is good evidence that it can be effective.[10]

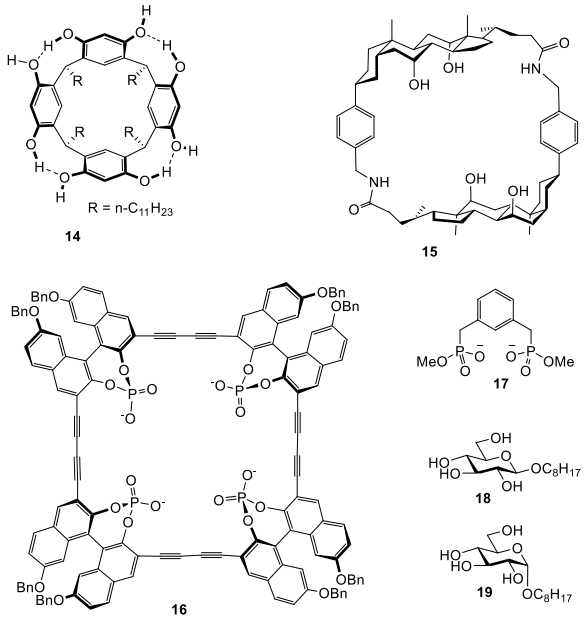


Fig. 3 First attempts: Early biomimetic carbohydrate receptors designed to operate in organic solvents, with standard substrates **18** and **19**.

The principle of full complementarity is well supported by crystallography of carbohydrate-binding proteins. Two examples are illustrated in Figs. 2d/e. The former shows fucose **4** in the binding site of a lectin from the fungus *Aleuria aurantia*.[11] A tryptophan provides hydrophobic/CH-π interactions to the upper face of the fucose, while other apolar amino acids help to form an apolar pocket for the methyl group. Five hydrogen bonds are formed involving arginine, aspartic and tryptophan side-chains. The fucose C1 and C2-OH groups are exposed to solvent and not directly involved in binding. K_a for the complex is 6000 M^{-1}, fairly typical for lectin-monosaccharide interactions. Fig. 2e shows β-D-glucose **1b** complexed to a monosaccharide-binding protein found in the periplasmic space of *E. coli*.[7] Axial CH groups again form hydrophobic/CH-π interactions to a tryptophan, with a phenylalanine performing the same role on the other side of the substrate. In this case the carbohydrate is fully enclosed, with ten hydrogen bonds to polar residues and two more to an immobilised water molecule. This is reflected in a K_a of 5×10^6 M^{-1}, among the largest known for protein monosaccharide interactions. The message appears to be that high affinities are possible, but only with optimum complementarity.

Finally, it was mentioned earlier that polar groups preorganised for carbohydrates may not be ideal for solvent water. A further possibility was raised by Lemieux – that the amphiphilic surfaces required to bind saccharides may interact surprisingly poorly with water, despite the presence of strongly polar groups.[12] This "hydraphobic" effect is difficult to discuss quantitatively, but could help to explain natural saccharide binding and would presumably assist supramolecular chemists. One might further speculate that the effect could be optimised, i.e. that binding sites could be designed both for complementarity to carbohydrates and, to some extent independently, for hostility to water. However, for the time being it is probably enough to match the target, and this is the focus of the remaining sections.

3. Carbohydrate Recognition in Organic Solvents

As discussed above, carbohydrate recognition in nonpolar media is intrinsically easier than in water, so it is unsurprising that most early work was conducted in organic solvents. The choice of substrate was not so trivial, as few carbohydrates are soluble in solvents such as chloroform, but glycosidic surfactants such as octyl glucosides **18** and **19** are suitable, readily available, and have been widely used. Progress up to 1999 was reviewed in some detail,[8] so only a few of the early examples will be mentioned here. Two of the first were **14**, first published in 1988 by Aoyama,[13] and **15**, described by the present author's group in 1990.[14] Both were studied with **18** and **19** in chloroform, using ^{1}H NMR titrations and, in the case of **14**, induced CD. Steroid-based **15** bound β-glucoside **18** with K_a = 1700 M^{-1},[14] and also showed modest (~3:1) enantioselectivity,[8] while resorcinarene **14** with **18** showed complex stoichiometry with somewhat lower overall affinity. Reported soon after were **16**, (due to Diederich; K_a to **18** ≈ 15,000 M^{-1} in CD_3CN:CD_3OD 98:2)[15] and **17** (due to Hamilton; K_a to **18** ≈ 26,000 M^{-1} in CD_3CN).[16] Receptors **14**-**16** illustrate the strategy depicted in Fig. 2a, i.e. surrounding the target with preorganised H-bonding functionality. On the other hand bis-phosphonate **17** demonstrates that sophisticated designs are not really needed. Even simple structures can achieve high affinities by using highly polar groups in a helpful medium.

Subsequent work developed both cyclic and acyclic approaches. Figure 4a shows a selection of the cyclic systems. Tricycle **20** was designed as a precursor for water-soluble "temple" receptors (see later) but the architecture also proved effective in non-polar solvents.[17] The geometry of the cavity is compatible with all-equatorial substrates such as β-glucoside **18**, and this target was bound

with K_a = 300,000 M^{-1} in chloroform. α-Glucoside **19** was complexed >20 times less strongly. A derivative related to **20** was able to extract glucose into chloroform from water.[18] Bicyclic cage **21**, due to Roelens, possesses a similar geometry and also bound **18** in chloroform (K_a ~50,000 M^{-1}).[19] In this case binding to α anomer **19** was undetectable. Martinez and Dutasta have studied a range of bicyclic cryptophanes, typified by **22**, with several octyl glycosides as substrates.[20] These receptors are less powerful (K_a up to 2500 M^{-1} in chloroform) but are readily varied (including sense of chirality) and show tuneable selectivities. Macrocycles **23**, due to Abe and Inouye, bind octyl glycosides with K_a up to 5×10^6 M^{-1} in DCE, but modest selectivity.[21]

Among acyclic systems, the 1,3,5-triethyl-2,4,6-tris(aminomethyl)benzene scaffold has proved especially popular. This unit, which is also present in **21**, is nicely sized to encompass a monosaccharide. The parent triamine is highly accessible and can be derivatised in a variety of ways, which may include loss of C_3 symmetry. Figure 4b shows two examples. Bis-phenanthroline **24** is one of many that have been studied by Mazik,[9] and was shown to bind β-glucoside **18** with K_a >10^5 M^{-1} in $CDCl_3$, even when 5% DMSO was added. Hexa-amine **25** belongs to a series made by Roelens and colleagues.[5] This podand binds octyl α-D-mannoside **26** quite strongly (K_a ~10^4 M^{-1}) in CD_3CN, a fairly competitive solvent.[22] α-Mannosides are interesting targets as they appear on the surface of various pathogenic organisms, including HIV-1. Indeed, **26** was tested for inhibition of this virus and showed significant activity. This highlights an important point; while fully biomimetic carbohydrate receptors should operate in aqueous media, water-solubility is not a prerequisite for biological activity. Provided a receptor can work in the presence of water, perhaps across a phase boundary, it has the potential to be useful.

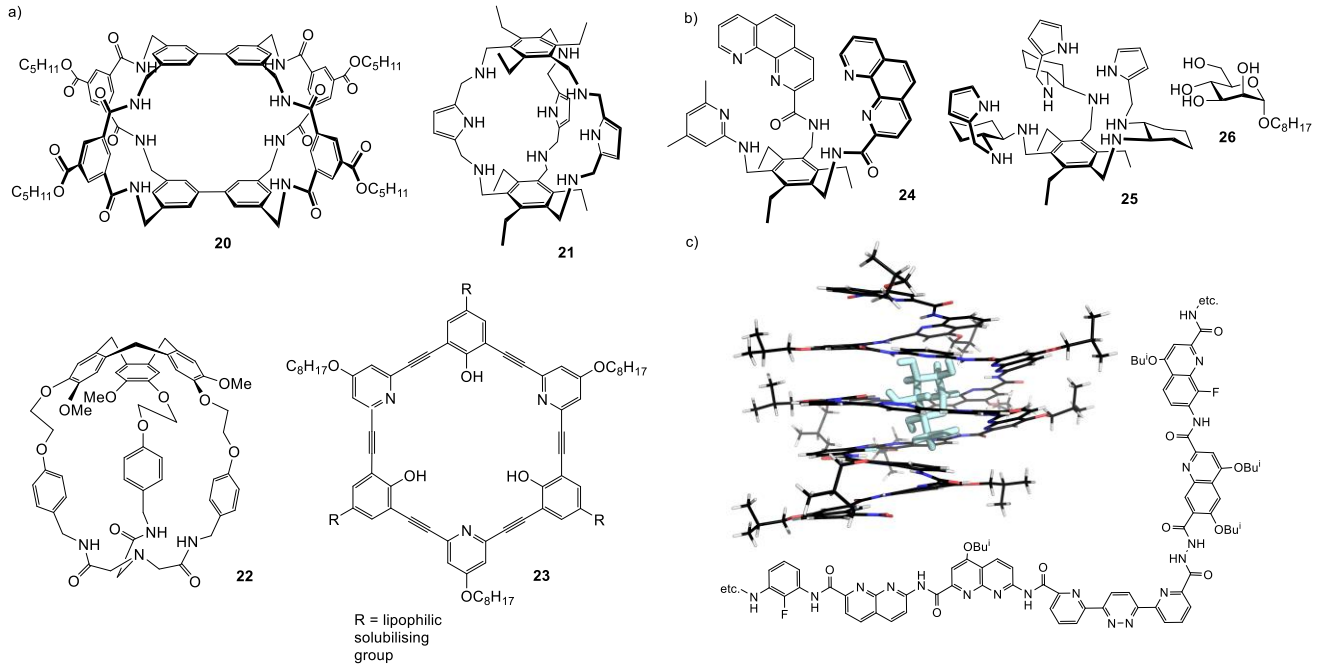


Fig. 4 Later generation receptors for carbohydrates in organic media. (a) Tri-, bi- and monocyclic systems **20**-**23**. (b) Podand-type receptors **24** and **25**, based on the 1,3,5-triethyl-2,4,6-tris(aminomethyl)benzene scaffold, and the octyl mannoside **26** targeted by **25**. (c) Crystal structure and partial formula of a linear foldameric receptor binding β-fructopyranose (cyan). The full receptor consists of 16 components linked through amide bonds. Six of the central units are represented in the formula.

Finally, Ferrand and Huc have developed a powerful approach employing linear oligomers which fold to create binding sites. A hexadecameric example binding a pyranose form of fructose is illustrated in Figure 4c.[23] This complex forms with K_a = 30,000 M^{-1} in 4:1 $CDCl_3$/[D_6]-DMSO, a highly competitive solvent system. A notable feature is the complete encapsulation of the saccharide, as revealed by X-ray crystallography (Fig. 4c). Access to the binding site is only possible through unwinding of the folded structure, in a process which mirrors the most powerful of the carbohydrate-binding proteins.[7] It would be interesting to see how well this system would perform if it could be translated to water, although this has not yet been achieved.

4. Biomimetic Carbohydrate Recognition in Water

4.1 Early approaches

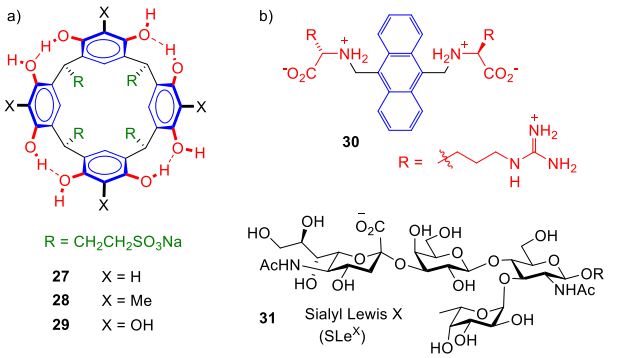


Fig. 5 Early receptors for carbohydrates in aqueous solution. (a) Water-soluble resorcinarenes **27**-**29**. (b) Nilsson's anthracene-arginine conjugate **30**, and substrate Sialyl Lewis X **31**. Aromatic surfaces potentially involved in hydrophobic/CH-π interactions are shown in blue, regions capable of polar interactions are coloured red, water-solubilising groups are green. This colour scheme is used for the remainder of the article.

Early attempts to bind carbohydrates through non-covalent interactions in water served largely to highlight the difficulty. Most work in the area involved the measurement of very low binding constants, or employed substrates which were relatively easy to attract from aqueous solution. Again, this research is well covered in previous reviews[8,24] so just two representative examples are described here. In line with Fig. 2b, both combine aromatic surfaces capable of hydrophobic/CH-π interactions with polar groups capable of hydrogen bonding and, where relevant, electrostatic interactions. Resorcinarenes **27**-**29** (Fig 5a), water-soluble relatives of **14**, were

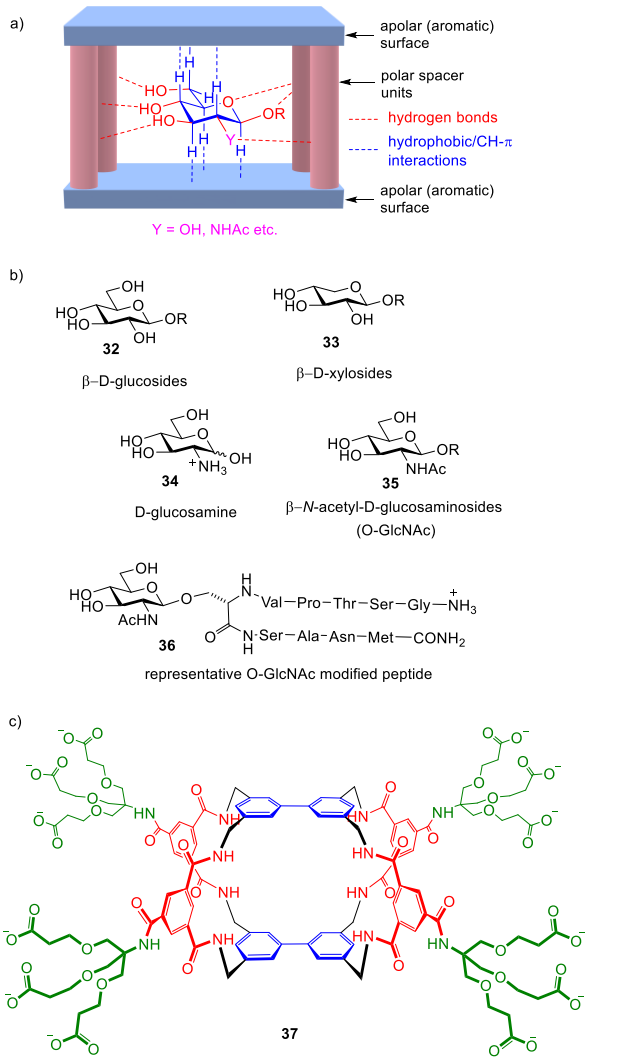


Fig. 6 (a) The "temple" approach to receptors for all-equatorial carbohydrates. (b) All equatorial monosaccharide targets (several more are shown in Fig. 1). (c) The first water-soluble temple.

reported by Aoyama in 1992. No binding was detected to glucose **1**, galactose **2** or mannose **3**, but fucose **4** formed weak complexes to the three macrocycles with K_a = 2, 6 and 8 M^{-1} respectively.[25] While fucose is not exactly an easy substrate, it is more hydrophobic than **1**-**3** and is presumably more readily separated from water. Higher affinities of ~100 M^{-1} were achieved by the acyclic receptor **30**, due to Nilsson, targeting important oligosaccharides such as Sialyl Lewis X (**31**).[26] However, despite their complexity, these substrates are probably less challenging than simple monosaccharides (provided selectivity is not required). Large targets present more surface area for non-covalent interactions, and the presence of functional groups other than hydroxyl (e.g. CO_2^-, NHAc in **31**) provides further opportunities for binding.

4.2 "Temple" receptors for all-equatorial carbohydrates

As recently as 2005, there were no well-characterised examples of biomimetic recognition in water targeting the common hexoses (glucose **1**, galactose **2**, mannose **3** etc.). A number of systems had been shown to work well for glycosides in organic solvents, but translation to aqueous solution had not been achieved. One example which had been fairly successful in non-polar media was tricyclic cage **20**, developed in the author's group. The rationale behind this design is shown in Figure 6a. The intention was to target the all-equatorial family of monosaccharides, e.g. glucose **1** and β-glucosides **32**; xylose **8** and β-xylosides **33**; glucosamine **34**, *N*-acetylglucosamine (GlcNAc) **5** and β-O-GlcNAc units **35**, especially in modified peptides and proteins such as **36** (see below). These substrates possess small, roughly equivalent patches of CH groups on both faces, potentially addressable by parallel aromatic surfaces capable of hydrophobic/CH-π interactions. The aromatic surfaces could be held apart by rigid polar spacers able to hydrogen bond to the substrate's equatorial polar functionality. Modelling suggested that biphenyl roof and floor units, combined with isophthalamide spacers, would generate a cavity with roughly the right dimensions for a monosaccharide. A cartoon depiction of the strategy suggested a classical temple (Fig. 6a), lending its name to the approach. In the case of **20** the biphenyl surfaces might not contribute greatly to binding in chloroform, but the intention was always to transfer the design to water by adjusting the externally-directed groups. In aqueous solution the weak CH-π interactions would be supplemented by hydrophobic effects, realising the benefits of full complementarity.

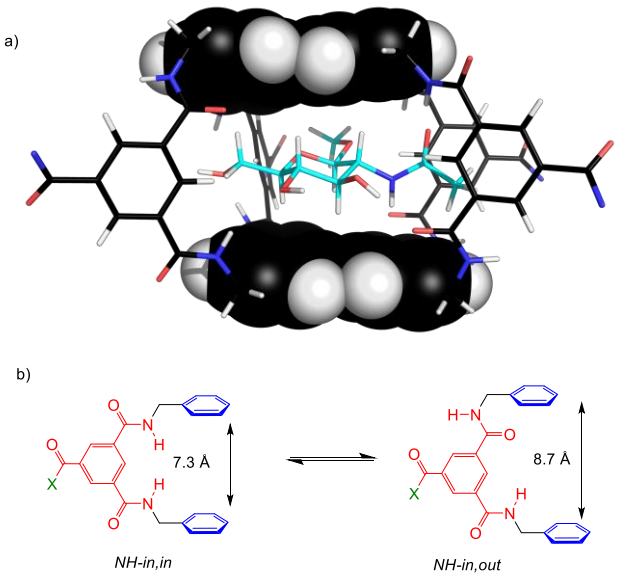


Fig. 7 (a) The structure of **37** complexed to GlcNAc-β-OMe (**35**, R = Me) in water, as determined by NOESY. (b) Conformational options of the isophthalamide spacer. The structure in (a) contains one *NH-in,in* spacer (back left), the remainder adopting the *NH-in,out* conformation.

The first water-soluble temple receptor was **37** (Fig. 6c), reported in early 2005.[27] The tricarboxylate solubilising groups, derived from dendrimer chemistry, were sufficient to ensure that the cage was monomeric in water and readily studied by NMR. The affinity of **35** for glucose was low, at K_a = 9 M^{-1}, but selectivity was fair (e.g. 5:1 for glucose **1** *vs.* galactose **2**), and this was the first clear example of biomimetic recognition of these substrates in water. Moreover, it was later found that **37** performed far better with the β-O-GlcNAc unit **35**. This monosaccharide has special importance as a dynamic post-translational modification of proteins, which is thought to play roles in diabetes, cancer and neurodenegerative diseases.[28] Glycoside **35** (R = Me) was bound with K_a = 630 M^{-1}, and modified peptide **36** with K_a = 1100 M^{-1}.[29] Usefully, the complex between **37** and **35** (R = Me) formed with slow kinetics so that its ^{1}H NMR spectrum could be observed directly. A full assignment was possible, 48 NOESY signals could be measured, and the structure shown in Figure 7 could be deduced. A significant feature is the variable geometry of the isophthaloyl spacers. Although these are fairly rigid, rotation around C-CO and N-CH_2 bonds is possible and this allows the cavity to breathe. As shown in Figure 7, the orientation with both NH groups directed inwards places the aromatic surfaces 7.3 Å apart, while rotating one NH outwards increases the spacing to ~8.7 Å. In the complex, three of the spacers adopt the *NH-in,out* conformation while just one exists as *NH-in,in*.

Meanwhile, the biphenyl-based design of **37** was developed in two ways; firstly by adding substituents to moderate binding properties (as in **38**), and secondly by extending the roof/floor units to terphenyls so that disaccharides could be addressed (as in **39** and **40**) (see Figure 8). In receptors **38**, a series of groups were added to the biphenyl 4,4'-positions, moderating the properties of the temple roof/floor without substantially changing the conformations.[30,31] CH-π interactions are greater for electron-rich π-surfaces so, for example, it was expected that **38** (Y = F) would be less effective than **37** or than **38** (Y = OR). In the event all variants of **38** bound glucose more strongly than **37**, suggesting that electronic effects are not so easy to manipulate.[31] On the other hand the highest affinity of 60 M^{-1}, for **38** (Y = OPr), represented a useful improvement on **37**.[30] In tetracycle **39**, *m*-terphenyl surfaces were used to create a binding site for cellobiose **10** and other all-equatorial disaccharides.[32] Receptor **39** bound cellobiose with K_a = 600 M^{-1} and 50-fold selectivity vs. lactose **11**, a disaccharide with just one axial OH. The design of **39** was based on the notion that five spacers would be required to maintain an open cavity. Modelling of the tricyclic alternative **40** (R = H) had suggested that the cavity should close via a twisting motion. When receptors **40** were prepared somewhat later, the results were unexpected; affinities for cellobiose **10** were raised to 3000 M^{-1}, despite the simpler, less connected framework.[33] The success of these receptors highlights the danger of trusting too much in modelling, especially when predictions are negative.

Although the bi/ter-phenyl units were clearly effective as hydrophobic surfaces, they are probably not ideal. The tendency for Ar-Ar bonds to twist means that extended planar structures are disfavoured, and this restricts the potential for simultaneous CH-π interactions as in Fig 2c. It is notable that biology often uses a condensed aromatic, the tryptophan side-chain, to connect with carbohydrate CH groups (see Figs. 2d,e). Condensed aromatics present large continuously planar surfaces which can allow substrates to move without losing contact. They also confer greater rigidity and predictability on macrocyclic architectures, and possess optical properties which may be useful in sensing.

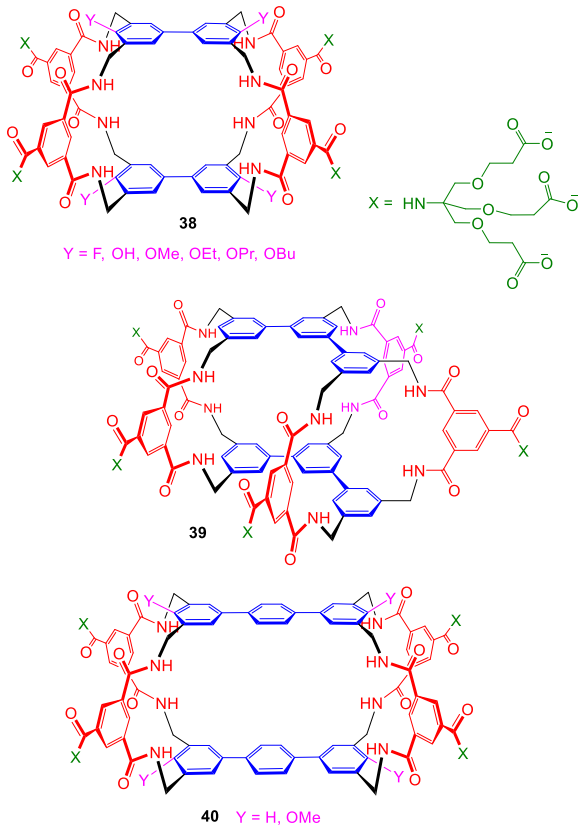


Fig. 8 Further bi- and ter-phenyl based temple receptors.

Accordingly, a second series of temple receptors was designed to exploit condensed aromatics as roof/floor units, starting with the simple monocyclic **41** (Fig. 9) This receptor illustrates the virtues of condensed aromatics quite well. The anthracene units are rigid in themselves, and also restrict movement through steric interactions between the amide nitrogens and peri-hydrogens (Fig. 10). As a result, the conformation of **41** is predictable with confidence, the only flexibility/uncertainty involving the angle between the anthracene planes. This preorganisation is achieved without a polycyclic architecture as employed in **38**-**40**. Receptor **41** was synthesised via a straightforward 2 + 2 cyclisation, and found to bind glucose with K_a = 56 M^{-1}.[34] Moreover the anthracene fluorescence increased 2.5-fold on binding, raising the possibility of glucose sensing (Fig. 10). An X-ray crystal structure of **41** in the presence of glucose confirmed the expected binding mode (Fig. 10);[35] this remains the only crystal structure of a fully biomimetic (i.e. water-compatible) carbohydrate receptor. The core of **41** was also employed to test the effect of varying the side-chains. A number of dendrimeric polycarboxylates were tested, and small enhancements in glucose binding were observed. For example, a 1.5-fold increase to 90 M^{-1} was observed for **42**, perhaps due to interactions between glucose and the side-chains.[36] Unsurprisingly, the systems with more carboxylates (up to 54) bound glucosammonium **34** quite strongly (up to 7000 M^{-1}).

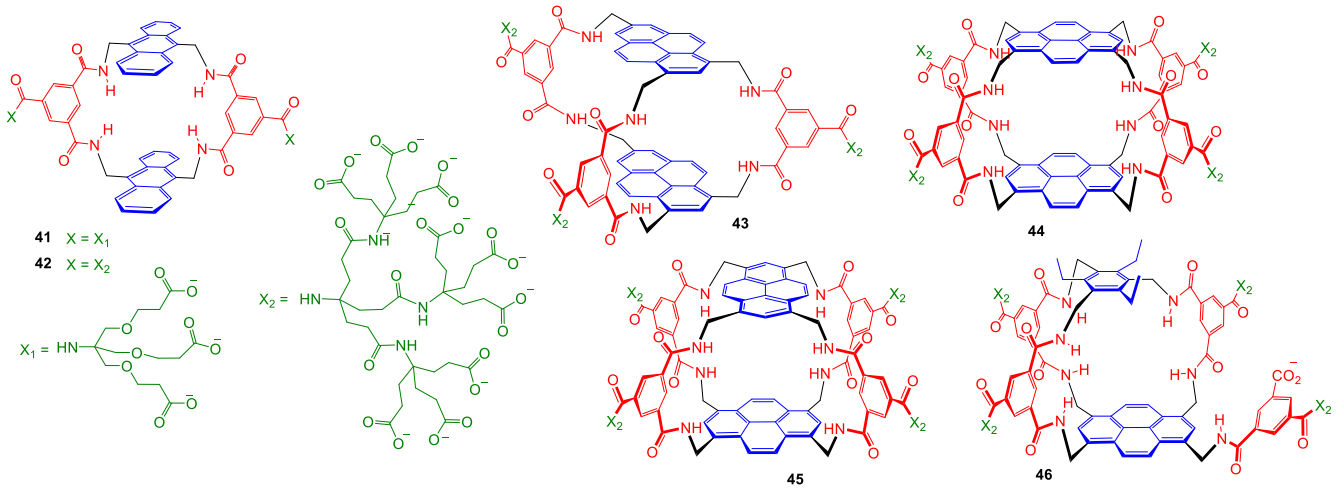


Fig. 9 Temple receptors based on condensed aromatic roof/floor units.

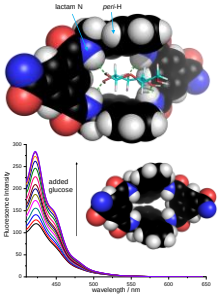
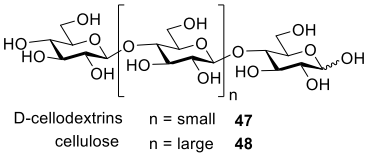


Fig. 10 Top: X-ray crystal structure of the complex between D-glucose and receptor **41**.[35] Side chains are omitted for clarity, intermolecular H-bonds are shown as green. Steric clashes between lactam N and *peri*-H atoms help to control flexibility. Bottom: Model of receptor **41** in absence of glucose, and variation in fluorescence emission as glucose is added. The empty receptor retains an open cavity although the anthracenes can tilt so that they meet at one end. When the glucose enters the cavity the angle between the anthracenes changes, presumably causing the increase in fluorescence.



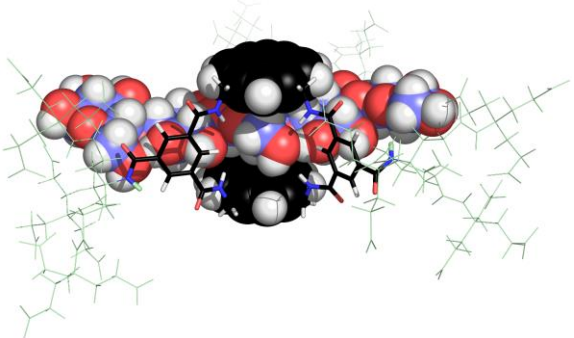


Fig. 11 Oligo- and polysaccharide substrates for receptor **43**, and model of the pseudorotaxane formed from **43** and cellopentaose **47** (n = 3).

Given the success of biphenyl surfaces in **37** and **38**, mutation of these units to pyrenes was a natural development. The pyrene unit is geometrically similar to biphenyl, but rigidified and with a larger surface area. The first pyrene-based temple was bicyclic **43**, chosen in part because a straightforward (if lengthy) directed synthesis was feasible.[37] Dispersing the pyrene surfaces in water was challenging, and the second-generation dendrimer from **42** was required for full solubility. Performance with monosaccharides was modest (e.g. K_a to glucose = 120 M^{-1}), but better results were obtained with all-equatorial oligosaccharides such as the cellodextrins

Table 1. Binding of glucose to temple monosaccharide receptors; association constants (K_a) and selectivities *vs.* galactose where available.

Receptor	K_a (M^{-1}) to D-glucose in water	Selectivity vs. D-galactose
37	9	4.5
38 (R = OPr)	60	20
41	56	14
42	90	13
43	120	7
44	120	
45	190	
46[a]	250	

[a] Unidentified enantiomer

(cellulose fragments) **47** (Fig 11). For example cellotetraose **47** (n = 2) was bound with K_a = 12,000 M^{-1}.[37] The portals in **43** are wider than in previous temples, and NMR studies supported the formation of threaded pseudorotaxane complexes with the oligosaccharides (Fig. 11). Moreover, AFM evidence was obtained for binding of **43** to cellulose **48**, forming multiply threaded polypseudorotaxane complexes. This raises the possibility that threading receptors like **43** could be used to modify the properties of cellulose, for example promoting solubility in water. Cellulose is the most abundant organic material on earth, so any development which helps its exploitation could be very valuable.

Tricyclic cage **44** is directly analogous to prototype temple **37**, and might seem especially promising as a monosaccharide receptor. The directed synthesis of **44** proved unrealistic but, by creating all three rings in a single step reaction between **49** and **50** (Fig. 12), it was possible to prepare both **44** and isomer **45** in acceptable yields.[38] The isomers were challenging to separate and identify, but both had interesting binding properties. Affinities to glucose were moderate (120 and 190 M^{-1} for **44** and **45** respectively), but binding to the β-O-GlcNAc unit **35** was promising. For example tricycle **45** bound GlcNAc-β-OMe **35** (R = Me) with K_a ~18,000 M^{-1}, while the glycopeptide **36** (Fig. 6) was complexed to **44** with K_a = 67,000 M^{-1}. These affinities are high enough to suggest that applications in glycobiology, for example the detection of O-GlcNAc in specific protein environments, may soon be feasible. Finally, intermediate **49** was also combined with triamine **51** (Fig. 12) to give the chiral bicyclic receptors ±-**46**.[39] Although the enantiomers could not be separated, they could be studied independently in the racemate. This system showed the highest affinities yet for simple underivatised monosaccharides; glucose **1** and GlcNAc **5** were bound with K_a = 250 and 1280 M^{-1} respectively (in each case by one unidentified enantiomer). In the case of GlcNAc **5**, the affinity for the second enantiomer could also be measured, at 80 M^{-1}, implying 16:1 enantioselectivity.

In summary, a range of temple receptors were reported between 2005 and 2017, featuring various roof/floor units separated by the ubiquitous isophthalamide spacer groups. They were not all targeted at glucose, and indeed some performed quite well in other respects. However, for those with glucose-sized cavities it is interesting to survey their performance against this important substrate. Table 1 gives a list of binding constants, as well as selectivities *vs.* galactose, a benchmark competitor. There is a clear sense of progression as affinities rise from very low to a level which might be termed respectable, considering the challenge. Although the culminating value of 250 M^{-1} is still very small by general biological standards, it is worth noting that common lectins are not so far

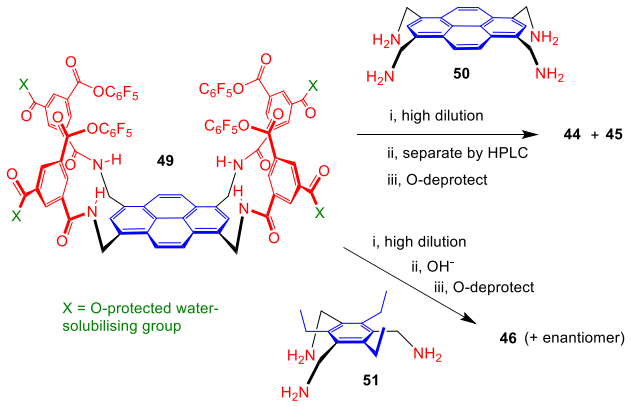


Fig. 12 Key synthetic steps leading to temple receptors **44** – **46**.

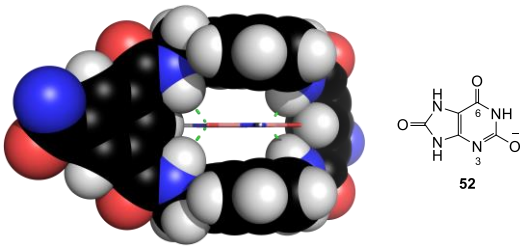


Fig. 13 Model of uric acid monoanion **52** bound to receptor **41**. Receptor side-chains are omitted for clarity. The four receptor NH groups form hydrogen bonds to the urate N3 and 6CO. The anthracene surfaces are nearly parallel, in contrast to the glucose complex in Fig. 10, implying a good fit between host and guest.

ahead. For example the lectin used to bind glucose, Concanavalin A, shows K_a of just 520 M^{-1},[6] around double the synthetic system. For GlcNAc the temples' performance is arguably quite impressive. The standard lectin for β-GlcNAc is Wheat Germ Agglutinin, which binds **35** (R = Me) with K_a = 730 M^{-1}.[38] In comparison, the affinity of **45** for the same substrate is 18,000 M^{-1}, twenty-five times higher.

This said, if the aim is to create useful carbohydrate binding agents, these temples are not quite satisfactory. For most applications one would want higher affinities, and accessibility is a problem in many cases. Moreover there is a general issue which remains hidden during studies of carbohydrate recognition, but becomes apparent when applications are considered. To illustrate, we return to the anthracene-based monocyclic receptors such as **41**. Of all the systems described above, these appear to have the most real-world potential. They are easy to synthesise, feature optical reporting, and bind glucose with affinities which are quite suitable for monitoring in biological samples such as blood. Blood glucose levels are around 6 mM, and more powerful receptors would be saturated and insensitive to variations around this concentration. Unfortunately, while these molecules can be used to monitor glucose in water, they give no signal variation in blood serum. They clearly bind something else and, considering the dimensions of the isophthalamide spacer (Fig. 7b), this may not be surprising. In the *NH-in,in* conformation, which pertains in **41**, the aromatic surfaces are 7.3 Å apart. This seems to be slightly too small for a carbohydrate. In the NMR structure of **37**·GlcNAc-β-OMe (Fig. 7a) most spacers adopted the wider *NH-in,out* conformation, while in **41**·glucose (Fig. 10) the guest sits towards one end of the cavity, pushing the anthracene surfaces apart and seemingly unable to reach the centre. However, 7.3 Å is very close to the inter-base spacing in DNA and almost ideal for an aromatic substrate. A polar aromatic which can form H-bonds to the lactams is potentially an excellent substrate, and a number of such molecules do in fact bind the bis-anthracene macrocycles with K_a as high as 10^7 M^{-1}. Uric acid, present as monoanion **52**, occurs in blood at quite high levels and is probably the main interferent (see Fig. 13). The isophthalamide spacer is ubiquitous in all the temples discussed thus far, so this issue is likely to be general whether or not it has been investigated. Binding saccharides in water is difficult enough, but it is still more challenging to ignore everything else that might be present.

Although these systems may not find practical applications, they have contributed theoretically. Unlike proteins, synthetic receptors can be used in a wide range of solvents, especially when peripheral modifications can be made. This allows studies of solvent effects, which can throw light on the driving force for binding. Receptors **37** and **39** were studied with glucose **1** and cellobiose **10** respectively, in a series of polar solvent mixtures (water + MeOH/DMSO/MeCN).[40] The corresponding organic-soluble precursor macrocycles were also investigated with octyl glycosides in chloroform + MeOH mixtures. In the latter experiments, H-bonding is dominant and the addition of methanol to the chloroform (increasing solvent polarity) reduced affinities as expected. However, in the polar solvent mixtures, addition of organic solvents to water (*decreasing* polarity) also depressed binding. This implies that water is less comfortable than the organic solvents in the amphiphilic receptor cavities, and that hydrophobic/hydraphobic effects do make a contribution to binding. If **37** and **39** are accepted as models of proteins, this conclusion extends to natural carbohydrate recognition and helps to elucidate an important biological phenomenon.

4.3 Alternative solutions – platforms, strands and toroids

The temple architecture is limited in substrate scope, and different approaches will be needed to address the full range of saccharide structures. A number of alternatives have appeared over the past decade or so. One popular strategy is to employ a central aromatic platform with potential for apolar interactions, surrounded by polar groups which can hydrogen bond to substrates and/or confer water solubility. With just one apolar binding surface and no "roof", such structures should be compatible with substrates bearing axial substituents, and could be quite versatile. Four systems conforming to this approach are shown in Figure 14. The simple benzene derivative **53**, due to Mazik, was only studied with all-equatorial substrates, methyl β-D-glucoside **32** (R = Me) and cellobiose **10**, but gave encouraging results with the latter.[41] Analysis of ^{1}H NMR data suggested two binding events with successive K_a values of 305 and 66 M^{-1}. In **54**, from the group of Meldal, the peripheral polar groups consist of a cyclic peptide creating a nicely preorganised binding site. Unfortunately complex formation could only be quantified by 2D ^{1}H NMR titrations, and only weak binding (8 M^{-1}) to cellobiose **10** was reported.[42] Receptor **55**, a derivative of "Mallard Blue", bears some resemblance to Nilsson's anthracene **30**. As in that case, a simple and accessible structure is used to target a large and complex substrate, in this case the charge-complementary polysaccharide heparin **56**.[43] **55** binds heparin quite strongly, signalling through changes to its UV-visible spectrum, and has potential for medically relevant sensing applications. Most recently, the pyrenes **57** and **58** were designed to

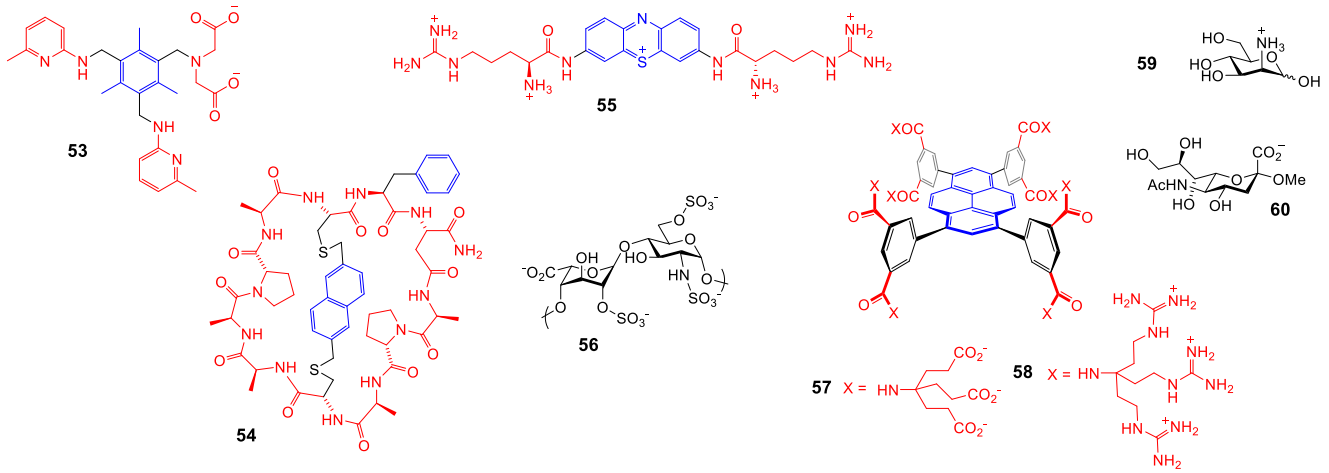


Fig. 14 Platform-type designs for carbohydrate receptors, with relevant substrates.

present two identical binding sites, aimed primarily at substrates with charged axial groups. Accordingly, anionic **57** bound mannosamine **59**, and cationic **58** bound sialoside **60**, with first (1:1) K_a values of 3000 M^{-1} and 1300 M^{-1} respectively.[44] Complexation on both faces was detected, mimicking the multivalency shown by many lectins.

A second approach involves linear oligomeric structures containing aromatic and polar units. Three examples are shown in Figure 15. Nonaphenyl **61** forms a homo-double helix in water, but in the presence of heptasaccharides this unwinds and binding to the carbohydrate may be detected by CD. However, the effect is difficult to quantify because of complex stoichiometries.[45] The phenol-pyridine oligomer **62**, an acyclic water-soluble relative of **23**, showed evidence of very weak binding to glucose, but was more successful with aminosugars; its affinity for glucosamine **34** was measured as K_a = 2000 M^{-1}.[46] Both **61** and **62** are composed of rigid units so that conformational freedom is limited. It may be difficult to predict exactly how they will fold, but the chances of clefts or cavities may be relatively high. The third example, peptide **63**, is more flexible and emerged from a different concept. In a study of borylated peptides as oligosaccharide receptors, Hall and coworkers made a library of pentapeptides each containing two -B(OR)$_2$ groups, and screened for binding to the Thomsen-Friedenreich antigen **64**.[47] Disaccharide **64** is a tumour-associated antigen and so an important target. A diborylated analogue of **63** emerged from the screening and was found to bind **64** with K_a = 2000 M^{-1} (presumably in non-biomimetic fashion, involving B-O bonds). More importantly, from the viewpoint of this article, **63** itself was prepared as a control and retained much of the affinity; K_a was lowered, but only as far as 400 M^{-1}. The result suggests that short peptides with aromatic components could be quite effective as biomimetic carbohydrate receptors, and might be discoverable by combinatorial methods.

Macrocyclic, toroidal structures have also been successful in some cases. The coordination capsule **65** (Fig. 16), due to Yoshizawa, possesses an apolar interior composed of extended aromatic surfaces.[48] Although this might not seem promising for carbohydrate binding, in that some polar interactions would usually be expected, macrocycle **65** is an effective and selective receptor for sucrose **66**. The affinity for sucrose was measured at 1170 M^{-1}, and the complex was formed in the presence of several disaccharide competitors (e.g. cellobiose **10**, lactose **11**, maltose **12**). Although it is interesting that such a hydrophobic cavity can succeed, sucrose is probably a special case, capable of adopting a conformation with a substantially apolar exterior. Cucurbiturils such as **67** also possess hydrophobic interiors, in this case supplemented by externally directed polar carbonyl oxygens. The group of Kim have shown that **67** is a good receptor for aminosugars, binding protonated glucosamine **34**, galactosamine **68** and mannosamine **59** with impressive affinities of 4400, 16000 and 1900 M^{-1}.[49] At first sight these results may seem surprising, as host and guests are not

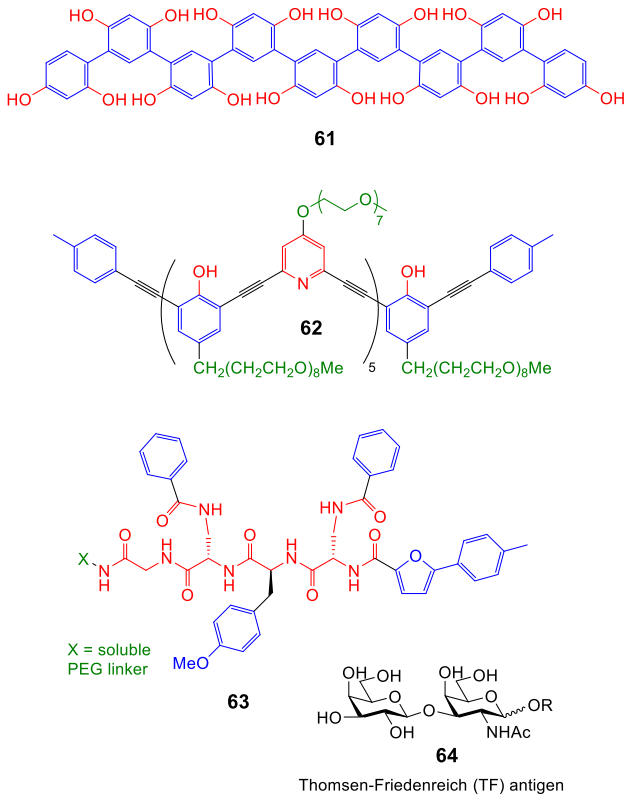


Fig. 15 Linear oligomeric receptors for carbohydrate recognition in water.

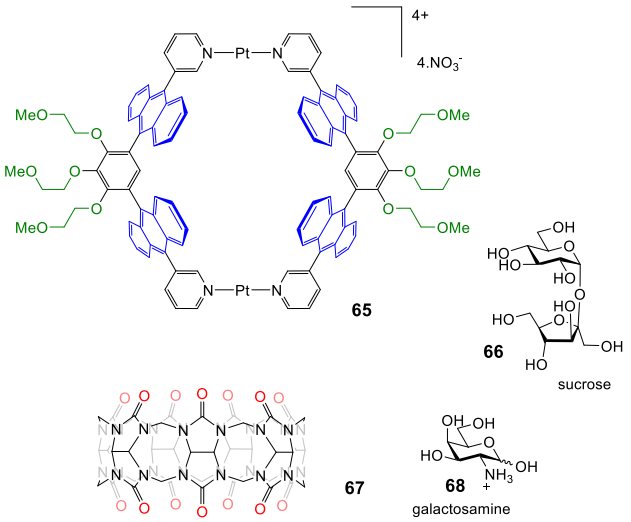


Fig. 16 Macrocyclic carbohydrate receptors.

obviously complementary. However, they are less remarkable when one considers that other ammonium cations are bound far more strongly; for example cyclo-octylammonium with $K_a = 3 \times 10^{11}$ M^{-1}. The affinities for aminosugars probably reflect the extreme hydrophobicity of **67** more than its compatibility with carbohydrates.

Finally the macrocycle **69**, due to Francesconi, Roelens, Nativi and colleagues, possesses an amphiphilic cavity with anthracenyl aromatic surfaces and diaminocarbazoles as polar spacers.[50] This receptor bears similarities to the temple family discussed in Section

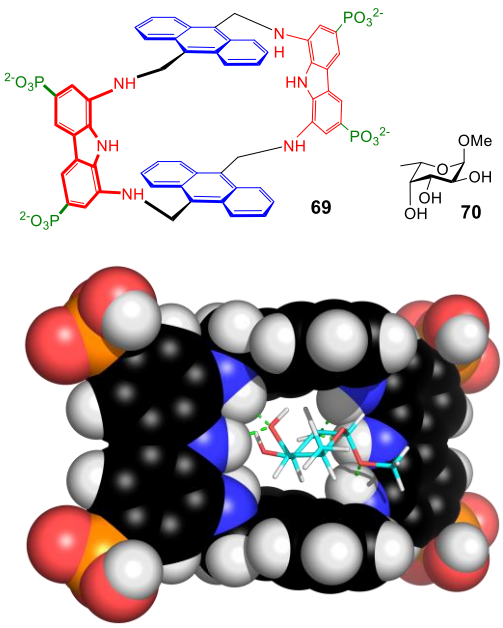


Fig. 17 Anthracene-carbazole macrocycle **69,** methyl α-L-fucoside **70**, and proposed structure of **69**·**70**. **69** is modelled as the neutral tetraphosphonic acid. Five NH···O hydrogen bonds, between 1.9 and 2.1 Å, are formed between host and guest. The spacing between anthracene units is ~8 Å. We thank Dr. O. Francesconi for providing the coordinates of the modelled complex.

4.2, especially the bis-anthracenes **41** and **42**. Its behaviour is complicated by dimerisation leading to multiple binding stoichiometries, but could be quantified using "median binding concentration" (BC^0_{50}), a parameter designed by the authors to reflect overall binding ability. For all-equatorial methyl β-D-glucoside **32** (R = Me), BC^0_{50} values equivalent to K_a ~1000 M^{-1} were measured, comparing well with the temples discussed in section 4.2. Interestingly, however, the highest affinities of K_a ~2000 M^{-1} were measured for methyl α-L-fucoside **70**, with two axial groups. Molecular modelling guided by NOESY suggested the structure for **69**·**70** shown in Figure 17. Complex formation appears to be driven by formation of five H-bonds as well as hydrophobic/CH-π interactions. The distance between anthracene units is ~8 Å, slightly larger than for **41**, and this may be advantageous (see next section). The structure also shows how some axial substitutions may be compatible with parallel aromatic surfaces, suggesting that temple and related architectures may be more versatile than originally supposed.

4.4 The Hexaurea Temple – hitting the sweet spot for glucose

As described in Section 4.2, the temple architecture was designed to address all-equatorial targets, of which the most important is glucose. On the positive side, moderate affinities (up to 250 M^{-1}) were achieved, enough for some applications, while selectivities against most other carbohydrates were fairly good (see Table 1). Unfortunately the most promising candidate **41** suffered from strong binding to non-carbohydrate substrates, and it seemed likely that this issue would affect the whole family of temple receptors. The problem lies with the isophthalamide spacer, which allows a spacing of 7.3 Å between roof and floor surfaces (Fig 7b), and in some designs (e.g. **41**) enforces this distance. 7.3 Å is excellent for binding aromatics, but at least 1 Å too small for carbohydrates.

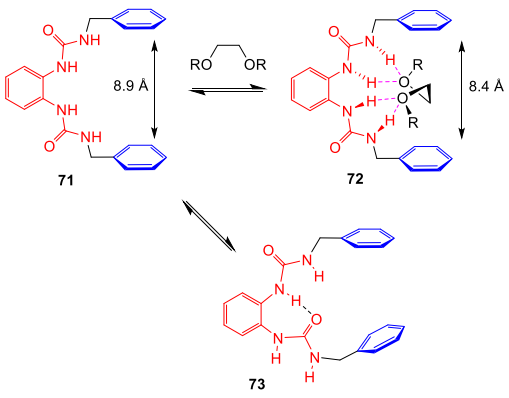


Fig. 18 *o*-Phenylene-bis-urea, an alternative spacer for temple receptors.

Considering alternatives which might expand the cavity, the bis-urea unit **71** seemed interesting (Fig. 18). One bond longer than the isophthalamide, it places the aromatic surfaces up to 8.9 Å apart. When presented with vicinal OR groups it is able to make four hydrogen bonds, as in **72**, while the spacing contracts slightly to ~8.4 Å. Disadvantages are its ability to access conformation **73** with an intramolecular H-bond, and the tendency of oligoureas to insolubility. However, its potential is especially clear when it is used to separate 1,3,5-trisubstituted benzene units. Although such small roof/floor components could not be used with the isophthalamide spacers, the extra length of the bis-ureas allows shrinkage of the apolar surfaces while maintaining cavity volume. Hexaurea **74**[51] (Fig. 19), designed to prototype this new family of temples, serves to illustrate its advantages. The interior of **74** can accept a monosaccharide and, being C_3 symmetric, is roughly congruent with a six-membered pyranose ring. When β-D-glucose is placed inside the cavity, the complementarity is remarkable (Fig. 19). All the glucose oxygens bar one (the anomeric OH), and all the ureas bar one, are involved in intermolecular hydrogen bonds, and all the axial CH make contact with the cavity roof/floor. Altogether there are ten H-bonds and five CH-π interactions to drive complex formation. In the absence of glucose, modelling indicates that the geometry of the cage changes very little, maintaining an open cavity. The contrast with earlier temples (Section 4.2) is stark. These systems provided plenty of π-surface for apolar interactions, but polar interactions were essentially left to chance and certainly not optimal. The hexaurea temple was the first to feature a rationally conceived, almost comprehensive pattern of polar interactions.

In addition to the core hexaurea cage, **74** incorporates two further design elements with important roles. Three nonacarboxylate solubilising groups, previously used for many of the earlier temples, guarantee water-solubility, and six ethyl groups constrain the conformational freedom of the core and promote an open binding site. Both these features also assist with the synthesis, shown in Figure 20.[51] Triethyltriamine **51** is readily available,[52] and has featured in several earlier carbohydrate receptors.[5,19,9,22,39] Its derivatives favour an up-down alternating conformation which presumably assists the last step. The solubilising groups are O-protected until the final step and help to maintain solubility in organic media throughout the synthesis. It is notable that yields in the cyclisation step were greatly improved by addition of octyl β-D-glucoside **18**. The synthesis in Fig. 20 is remarkably short, especially in comparison to some of the earlier temples discussed in Section 4.2.

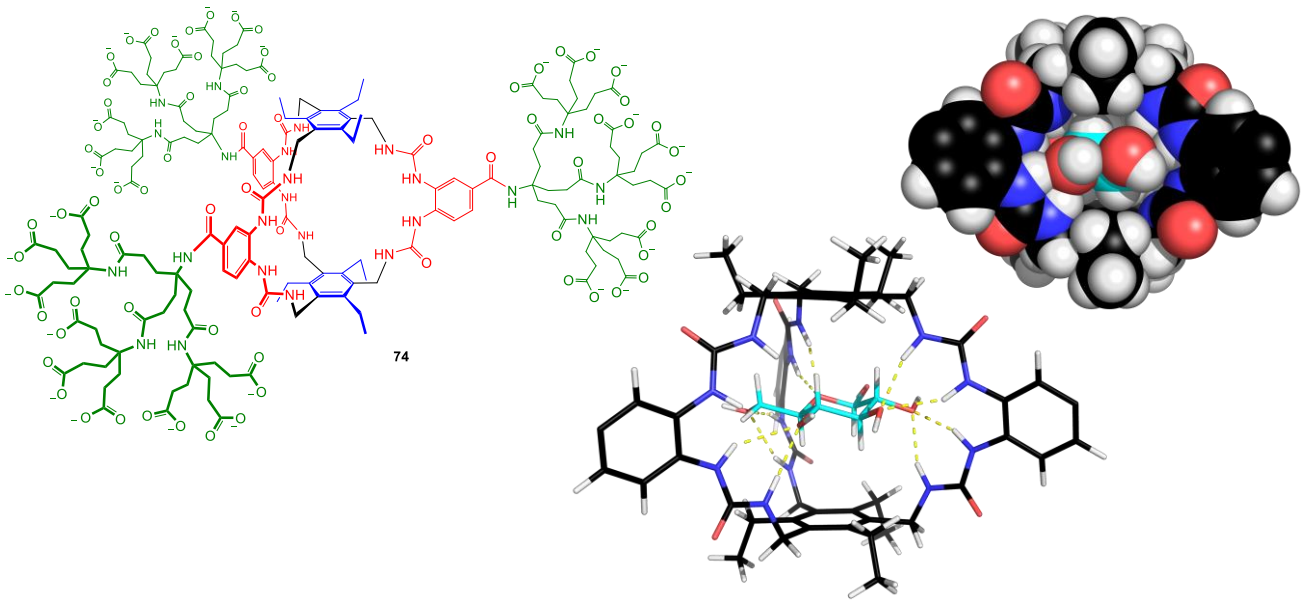


Fig. 19 Hexaurea temple **74**, and two representations of its complex with β-D-glucose **1b** (side-chains omitted for clarity). Ten hydrogen bonds (1.95 – 2.48 Å) are formed between host and guest. The spacing between roof/floor benzene units is ~8.4 Å.

Receptor **74** was expected to bind strongly and selectively to β-glucosyl units, and did not disappoint. Results from binding studies are summarised in Figure 21. Glucose itself was bound with K_a ~18,000 M^{-1} in aqueous solution, determined by both ^{1}H NMR and ITC. This represents a seventy-fold increase over the best of the earlier temples (see Table 1), and is higher than most lectin-monosaccharide interactions (cf. Fig. 2d).[6]. Only the very strongest natural glucose receptors show greater affinities (Fig. 2e).[7] Methyl β-D-glucoside **32** (R = Me) was also bound, with K_a = ~8000 M^{-1}; this is consistent with the model in Fig. 19 which places the anomeric OH in a portal, such that some modifications should be allowed. A few other all-equatorial substrates were also bound fairly well: xylose **8** at 5,800 M^{-1}; glucuronic acid **76** at 5,300 M^{-1}; 2-deoxyglucose **77** at 725 M^{-1} (though this represents a 25-fold decrease from glucose, resulting from loss of just one OH group). Among other carbohydrates, galactose **2**, mannose **3**, ribose **9**, fructose **78** and cellobiose **10** are bound to measurable extents (Fig. 21). Curiously, **74** is probably the strongest biomimetic galactose receptor yet reported (K_a = 180 M^{-1}), yet shows the highest selectivity for glucose *against* galactose (100:1).

Perhaps equally important is the list of compounds for which no binding could be detected. Based on ITC data for cellobiose (above) it is likely that affinities down to ~30 M^{-1} should have been measurable. Titrations were performed for several other carbohydrates (fucose **4**, *N*-acetylglucosamine **5**, methyl α-D-glucoside **79**, maltose **12**, ascorbic acid **80**), linear polyols (mannitol **81**, D-gluconic acid **82**) and a range of biologically relevant aromatic compounds (**83** – **89**, see Fig. 21). All gave negative results. Finally **74** was tested in serum, where it was found to bind glucose with K_a = 11,000 M^{-1}. Although this is slightly lower than in water, it should not

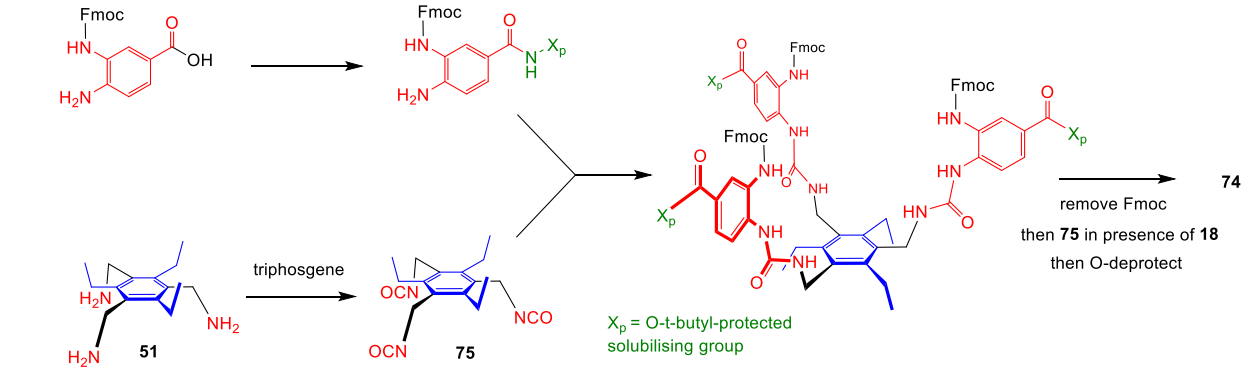


Fig. 20 Synthetic route to hexaurea temple **74**.

be enough to preclude applications in biological media.

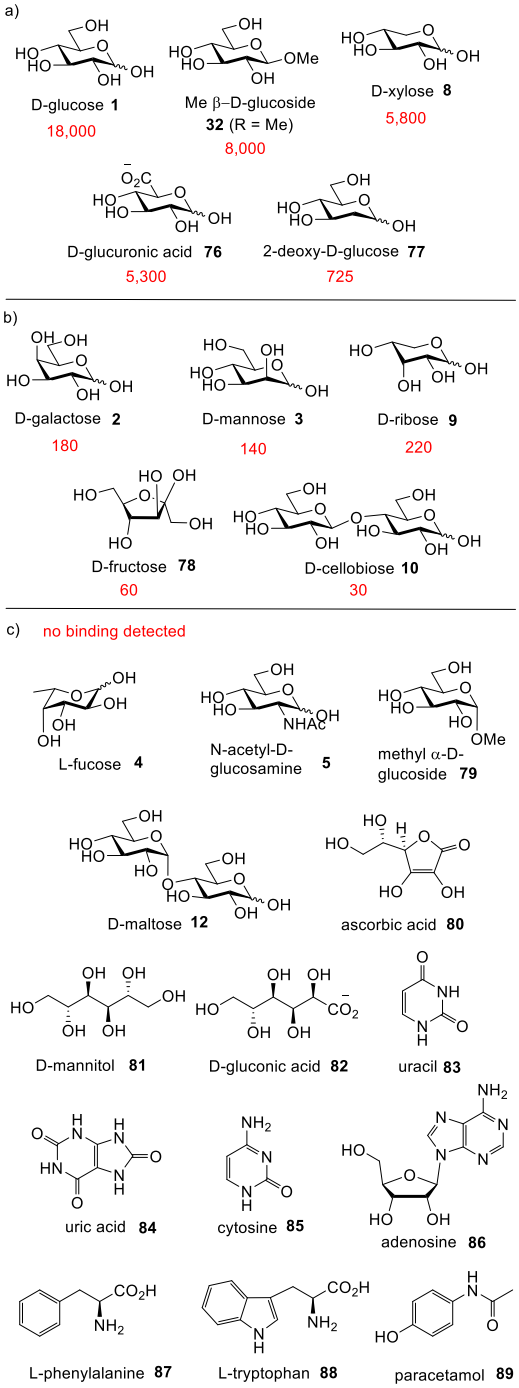


Fig. 21 Substrates and results for binding studies to **74** in 10 mM aqueous phosphate buffer (pH = 7.4). Numbers in red are binding constants (K_a, M^{-1}) obtained from ^1H NMR and/or ITC titrations. (a) All-equatorial substrates showing moderate to strong binding. (b)Weakly bound substrates. (c) Binding undetectable by ITC.

Clearly the hexaurea temple represents a major step forward for biomimetic carbohydrate recognition. At the outset of this research three decades ago, achieving any measurable binding to simple monosaccharides in water seemed enough of a challenge. Just a few years ago, an affinity of ~100 M^{-1} was considered significant (and was not always accompanied by good selectivity, especially against non-carbohydrates). Receptor **74** operates on a different level, with affinities and selectivities which might be considered high for proteins. Its selectivity is perhaps especially remarkable. According to SciFinder there are ~2,000,000 molecules with the same molecular weight as glucose, or less. Of these, the hexaurea temple binds glucose, xylose, deoxyglucose (less well) and probably a few other closely related structures (e.g. other deoxy- and fluorinated glucoses are likely substrates). There can be few other products of supramolecular chemistry which perform an important and difficult task so well.

5. Conclusions

The story told in this article ends on a positive note. At least for glucose, probably the most important substrate, we have a receptor that is fully biomimetic. The hexaurea temple **74** employs the same strategy and interactions as biology, works in biological environments, and performs to biological standards. Moreover, the prospects for real-world applications seem good. The receptor was commercialised through Ziylo, a company which had been spun out of the University of Bristol a few years previously. In August 2018 Ziylo (with rights to **74**) was sold to Novo Nordisk for a sum which could exceed $800 million depending on future developments. Novo Nordisk is the world's major supplier of insulin, and plans to exploit **74** in glucose-responsive variants of their products.[4] A new company, Carbometrics, was created to replace Ziylo in Bristol, and is collaborating with Novo Nordisk in this enterprise. At time of writing the research is still in its early stages, but the size of the deal attests to the confidence, of all concerned, that the venture can be successful. Meanwhile, Carbometrics has retained the right to exploit the hexaurea temple in non-therapeutic applications, and is also working towards this goal.

On the scientific side we can conclude, firstly, that the principles outlined in Fig 2 seem to be correct; complementing both polar and apolar substrate surfaces is the key to biomimetic carbohydrate recognition. This is no surprise given the evidence from protein structures, but it is welcome to find support from synthetic systems. Secondly, persistent efforts can lead to remarkably good performance. In at least one case there is a fairly simple and accessible receptor structure which works as well as we could hope. The question remains whether this will be an exception. We may find that other substrates are more difficult to accommodate, either because designs cannot be found or because promising structures are inaccessible. This said, optimal performance may not be needed for all applications, and useful affinities may be easier to achieve for some substrates. For example, tricycles **44** and **45** bind β-O-GlcNAc derivatives with K_a up to 70,000 M^{-1}, yet are surely not ideal for these targets.

Another issue is the role of combinatorial chemistry in carbohydrate recognition. For small monosaccharide targets, where a precisely defined cavity is needed, one feels that rational design will be more effective. However the balance may be different for larger substrates such as oligosaccharide antigens. Here it should be easier to achieve a useful degree of binding but more difficult to make rational predictions, especially as selectivity will be critical. Hall's discovery[47] that oligomer **63** binds disaccharide **64** suggests that a combinatorial approach could be successful. Libraries which incorporate rationally designed cavities and modular, variable regions may represent the best of both worlds, and it will be interesting to see if such strategies can be developed.

Conflicts of interest

The author was a director and shareholder of Ziylo at the time of the sale to Novo Nordisk, and is now a director and shareholder of Carbometrics.

Acknowledgements

The author thanks the many colleagues, co-workers and collaborators who contributed to research described in this article. Funding for the work has been provided by many organisations including Enterprise Ireland, the European Union, EPSRC, the Royal Society, and Novo Nordisk.

Notes and references

References

1. A guide into glycosciences: How chemistry, biochemistry and biology cooperate to crack the sugar code, D. Solis, N. V. Bovin, A. P. Davis, J. Jiménez-Barbero, A. Romero, R. Roy, K. Smetana and H. J. Gabius, *Biochim. Biophys. Acta-Gen. Subj.*, 2015, **1850**, 186-235.
2. The challenges of glycan recognition with natural and artificial receptors, S. Tommasone, F. Allabush, Y. K. Tagger, J. Norman, M. Kopf, J. H. R. Tucker and P. M. Mendes, *Chem. Soc. Rev.*, 2019, **48**, 5488-5505.
3. Glucose Sensing in Supramolecular Chemistry, X. L. Sun and T. D. James, *Chem. Rev.*, 2015, **115**, 8001-8037.
4. Glucose-responsive insulin by molecular and physical design, N. A. Bakh, A. B. Cortinas, M. A. Weiss, R. S. Langer, D. G. Anderson, Z. Gu, S. Dutta and M. S. Strano, *Nat Chem*, 2017, **9**, 937-943.
5. Biomimetic Carbohydrate-Binding Agents (CBAs): Binding Affinities and Biological Activities, O. Francesconi and S. Roelens, *ChemBioChem*, 2019, **20**, 1329-1346.
6. Structure and energetics of protein-carbohydrate complexes, E. J. Toone, *Curr. Opin. Struct. Biol.*, 1994, **4**, 719-728.
7. Protein-Carbohydrate interactions: basic molecular features, F. A. Quiocho, *Pure & Appl. Chem.*, 1989, **61**, 1293-1306.
8. Carbohydrate recognition through noncovalent interactions: A challenge for biomimetic and supramolecular chemistry, A. P. Davis and R. S. Wareham, *Angew. Chem., Int. Ed.*, 1999, **38**, 2978-2996.
9. Recent developments in the molecular recognition of carbohydrates by artificial receptors, M. Mazik, *RSC Advances*, 2012, **2**, 2630-2642.
10. Carbohydrate-Aromatic Interactions, J. L. Asensio, A. Arda, F. J. Canada and J. Jimenez-Barbero, *Acc. Chem. Res.*, 2013, **46**, 946-954.
11. Crystal structure of fungal lectin - Six-bladed beta-propeller fold and novel fucose recognition mode for Aleuria aurantia lectin, M. Wimmerova, E. Mitchell, J. F. Sanchez, C. Gautier and A. Imberty, *J. Biol. Chem.*, 2003, **278**, 27059-27067.
12. How water provides the impetus for molecular recognition in aqueous solution, R. U. Lemieux, *Acc. Chem. Res.*, 1996, **29**, 373-380.
13. Highly cooperative binding of alkyl glucopyranosides to the resorcinol cyclic tetramer due to intracomplex guest-guest hydrogen bonding: solvophobicity/solvophilicity control by an alkyl group of the geometry, stoichiometry, stereoselectivity, and cooperativity, Y. Kikuchi, Y. Tanaka, S. Sutaro, K. Kobayashi, H. Toi and Y. Aoyama, *J. Am. Chem. Soc.*, 1992, **114**, 10302-10306, and references cited therein.

[14]. Artificial Receptors for Carbohydrate Derivatives, R. P. Bonar-Law, A. P. Davis and B. A. Murray, *Angew. Chem., Int. Ed. Engl.*, 1990, **29**, 1407.

[15]. A new family of chiral binaphthyl-derived cyclophane receptors: complexation of pyranosides, S. Anderson, U. Neidlein, V. Gramlich and F. Diederich, *Angew. Chem., Int. Ed. Engl.*, 1995, **34**, 1596-1600.

[16]. Molecular recognition of carbohydrates: strong binding of alkyl glycosides by phosphonate derivatives, G. Das and A. D. Hamilton, *J. Am. Chem. Soc.*, 1994, **116**, 11139.

[17]. A tricyclic polyamide receptor for carbohydrates in organic media, A. P. Davis and R. S. Wareham, *Angew. Chem., Int. Ed.*, 1998, **37**, 2270-2273.

[18]. Phase Transfer of Monosaccharides Through Non-Covalent Interactions; Selective Extraction of Glucose by a Lipophilic Cage Receptor, T. J. Ryan, G. Lecollinet, T. Velasco and A. P. Davis, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 4863-4866.

[19]. A self-assembled pyrrolic cage receptor specifically recognizes beta-glucopyranosides, O. Francesconi, A. Ienco, G. Moneti, C. Nativi and S. Roelens, *Angew. Chem., Int. Ed.*, 2006, **45**, 6693-6696.

[20]. Emergence of Hemicryptophanes: From Synthesis to Applications for Recognition, Molecular Machines, and Supramolecular Catalysis, D. W. Zhang, A. Martinez and J. P. Dutasta, *Chem. Rev.*, 2017, **117**, 4900-4942.

[21]. D-3h-Symmetrical Shape-Persistent Macrocycles Consisting of Pyridine-Acetylene-Phenol Conjugates as an Efficient Host Architecture for Saccharide Recognition, H. Abe, T. Yoneda, Y. Ohishi and M. Inouye, *Chem. Eur. J.*, 2016, **22**, 18944-18952.

[22]. Antiviral Activity of Synthetic Aminopyrrolic Carbohydrate Binding Agents: Targeting the Glycans of Viral gp120 to Inhibit HIV Entry, O. Francesconi, C. Nativi, G. Gabrielli, I. De Simone, S. Noppen, J. Balzarini, S. Liekens and S. Roelens, *Chem. Eur. J.*, 2015, **21**, 10089-10093.

[23]. Iterative design of a helically folded aromatic oligoamide sequence for the selective encapsulation of fructose, N. Chandramouli, Y. Ferrand, G. Lautrette, B. Kauffmann, C. D. Mackereth, M. Laguerre, D. Dubreuil and I. Huc, *Nature Chem.*, 2015, **7**, 334-341.

[24]. Progress in biomimetic carbohydrate recognition, D. B. Walker, G. Joshi and A. P. Davis, *Cell. Mol. Life Sci.*, 2009, **66**, 3177-3191.

[25]. Complexation of hydrophobic sugars and nucleosides in water with tetrasulfonate derivatives of resorcinol cyclic tetramer having a polyhydroxy aromatic cavity: importance of guest-host CH-pi interactions, K. Kobayashi, Y. Asakawa, Y. Kato and Y. Aoyama, *J. Am. Chem. Soc.*, 1992, **114**, 10307-10313.

[26]. Amphiphilic anthracene-amino acid conjugates as simple carbohydrate receptors in water, J. Billing, H. Grundberg and U. J. Nilsson, *Supramol. Chem.*, 2002, **14**, 367-372.

[27]. Carbohydrate recognition in water by a tricyclic polyamide receptor, E. Klein, M. P. Crump and A. P. Davis, *Angew. Chem., Int. Ed.*, 2005, **44**, 298-302.

[28]. Protein O-GlcNAcylation: emerging mechanisms and functions, X. Y. Yang and K. V. Qian, *Nature Reviews Molecular Cell Biology*, 2017, **18**, 452-465.

[29]. A synthetic lectin for O-linked beta-N-acetylglucosamine, Y. Ferrand, E. Klein, N. P. Barwell, M. P. Crump, J. Jimenez-Barbero, C. Vicent, G. J. Boons, S. Ingale and A. P. Davis, *Angew. Chem., Int. Ed.*, 2009, **48**, 1775-1779.

[30]. A synthetic lectin for beta-glucosyl, N. P. Barwell, M. P. Crump and A. P. Davis, *Angew. Chem., Int. Ed.*, 2009, **48**, 7673-7676.

[31]. Substituent effects in synthetic lectins - exploring the role of CH-pi interactions in carbohydrate recognition, N. P. Barwell and A. P. Davis, *J. Org. Chem.*, 2011, **76**, 6548-6557.

[32]. A synthetic lectin analog for biomimetic disaccharide recognition, Y. Ferrand, M. P. Crump and A. P. Davis, *Science*, 2007, **318**, 619-622.

[33]. High-Affinity Disaccharide Binding by Tricyclic Synthetic Lectins, B. Sookcharoenpinyo, E. Klein, Y. Ferrand, D. B. Walker, P. R. Brotherhood, C. Ke, M. P. Crump and A. P. Davis, *Angew. Chem., Int. Ed.*, 2012, **51**, 4586-4590.

[34]. A simple and accessible synthetic lectin for glucose recognition and sensing, C. Ke, H. Destecroix, M. P. Crump and A. P. Davis, *Nature Chem.*, 2012, **4**, 718-723.

[35]. Crystal structure of a complex between beta-glucopyranose and a macrocyclic receptor with dendritic multicharged water solubilizing chains, P. K. Mandal, B. Kauffmann, H. Destecroix, Y. Ferrand, A. P. Davis and I. Huc, *Chem. Commun.*, 2016, **52**, 9355-9358.

[36]. Affinity Enhancement by Dendritic Side Chains in Synthetic Carbohydrate Receptors, H. Destecroix, C. M. Renney, T. J. Mooibroek, T. S. Carter, P. F. N. Stewart, M. P. Crump and A. P. Davis, *Angew. Chem., Int. Ed.*, 2015, **54**, 2057-2061.

[37]. A threading receptor for polysaccharides, T. J. Mooibroek, J. M. Casas-Solvas, R. L. Harniman, C. M. Renney, T. S. Carter, M. P. Crump and A. P. Davis, *Nature Chem.*, 2016, **8**, 69-74.

[38]. Synthetic Receptors for the High-Affinity Recognition of O-GlcNAc Derivatives, P. Rios, T. S. Carter, T. J. Mooibroek, M. P. Crump, M. Lisbjerg, M. Pittelkow, N. T. Supekar, G.-J. Boons and A. P. Davis, *Angew. Chem., Int. Ed.*, 2016, **55**, 3387-3392.

[39]. Enantioselective carbohydrate recognition by synthetic lectins in water, P. Rios, T. J. Mooibroek, T. S. Carter, C. Williams, M. R. Wilson, M. P. Crump and A. P. Davis, *Chem. Sci.*, 2017, **8**, 4056-4061.

[40]. Solvent effects in carbohydrate binding by synthetic receptors: Implications for the role of water in natural carbohydrate recognition, E. Klein, Y. Ferrand, N. P. Barwell and A. P. Davis, *Angew. Chem., Int. Ed.*, 2008, **47**, 2693-2696.

[41]. Carboxylate-based receptors for the recognition of carbohydrates in organic and aqueous media, M. Mazik and H. Cavga, *J. Org. Chem.*, 2006, **71**, 2957-2963.

[42]. Specific recognition of disaccharides in water by an artificial bicyclic carbohydrate receptor, T. Reenberg, N. Nyberg, J. O. Duus, J. L. J. van Dongen and M. Meldal, *Eur. J. Org. Chem.*, 2007, 5003-5009.

[43]. Mallard Blue: A High-Affinity Selective Heparin Sensor That Operates in Highly Competitive Media, S. M. Bromfield, A. Barnard, P. Posocco, M. Fermeglia, S. Pricl and D. K. Smith, *J. Am. Chem. Soc.*, 2013, **135**, 2911-2914.

[44]. Platform Synthetic Lectins for Divalent Carbohydrate Recognition in Water, T. S. Carter, T. J. Mooibroek, P. F. N. Stewart, M. P. Crump, M. C. Galan and A. P. Davis, *Angew. Chem., Int. Ed.*, 2016, **55**, 9311-9315.

[45]. Double helical oligoresorcinols specifically recognize oligosaccharides via heteroduplex formation through noncovalent interactions in water, H. Goto, Y. Furusho and E. Yashima, *J. Am. Chem. Soc.*, 2007, **129**, 9168-9174.

[46]. Saccharide Recognition and Helix Formation in Water with an Amphiphilic Pyridine-Phenol Alternating Oligomer, Y. Ohishi, H. Abe and M. Inouye, *Eur. J. Org. Chem.*, 2017, 6975-6979.

[47]. Synthesis, and screening of a library of peptidyl bis(boroxoles) as oligosaccharide receptors in water: Identification of a receptor for the tumor marker TF-antigen disaccharide, A. Pal, M. Berube and D. G. Hall, *Angew. Chem., Int. Ed.*, 2010, **49**, 1492-1495.

[48]. A polyaromatic nanocapsule as a sucrose receptor in water, M. Yamashina, M. Akita, T. Hasegawa, S. Hayashi and M. Yoshizawa, *Science Advances*, 2017, **3**, e1701126.

[49]. Cucurbit 7 uril: A High-Affinity Host for Encapsulation of Amino Saccharides and Supramolecular Stabilization of Their alpha-Anomers in Water, Y. Jang, R. Natarajan, Y. H. Ko and K. Kim, *Angew. Chem., Int. Ed.*, 2014, **53**, 1003-1007.

[50]. A Biomimetic Synthetic Receptor Selectively Recognising Fucose in Water, O. Francesconi, M. Martinucci, L. Badii, C. Nativi and S. Roelens, *Chem. Eur. J.*, 2018, **24**, 6828-6836.

[51]. A biomimetic receptor for glucose, R. A. Tromans, T. S. Carter, L. Chabanne, M. P. Crump, H. Li, J. V. Matlock, M. G. Orchard and A. P. Davis, *Nature Chem.*, 2019, **11**, 52-56.

[52]. 1,3,5-2,4,6-functionalized, facially segregated benzenes - Exploitation of sterically predisposed systems in supramolecular chemistry, G. Hennrich and E. V. Anslyn, *Chem. Eur. J.*, 2002, **8**, 2218-2224.